PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-308487
(43)Date of publication of application : 07.11.2000

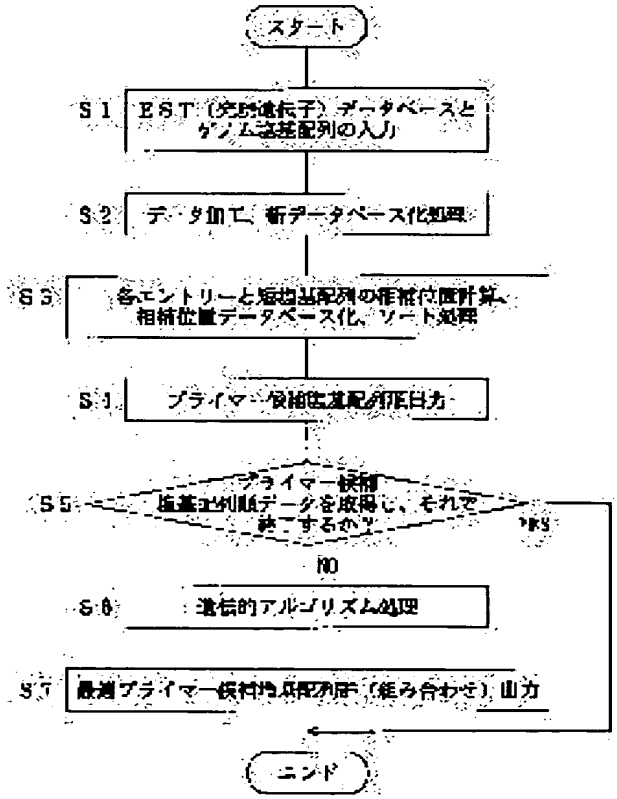
(51)Int.CI. C12N 15/09
C12Q 1/68
G06F 15/18
G06F 17/30

(21)Application number : 11-088410
(22)Date of filing : 30.03.1999

(71)Applicant : JAPAN SCIENCE & TECHNOLOGY CORP
(72)Inventor : MURAKAMI DAIYU
KANAZAWA ICHIRO
SUZUKI TAKASHI
GOTO JUN

(54) SELECTION OF PRIMER BASE SEQUENCE AND APPARATUS THEREFOR

(57)Abstract:
PROBLEM TO BE SOLVED: To obtain a method for selecting a primer base by obtaining a base sequence which is most suitable for the differential display method from an expression gene data base for processing, followed by selecting by the genetic algorithm using frequently appearing sequences as candidates.
SOLUTION: This is a method for selecting a primer base sequence by obtaining a base sequence which is most suitable for the differential display method from an expression gene data base for processing, using frequently appearing base sequences as candidates as an order of frequency, followed by selecting optimal primers from the obtained primer candidates by the genetic algorithm, by a method which make the differential display method more convenient which is an experimental method for the elucidation of the structure and function of a gene



and the molecular evolution, and for obtaining new genes. An expression gene data base is selected which was supplemented by overlapping an identical part by screening a homology to a genomic base sequence when the sequence has less than 1,000 bases.

LEGAL STATUS

[Date of request for examination]	19.03.2002
[Date of sending the examiner's decision of rejection]	03.08.2004
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	
[Date of registration]	
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

* NOTICES *

CLAIMS

[Claim(s)]
[Claim 1] The selection approach of the primer base sequence characterized by to give the process which selects the optimal primer group characterized by use of a genetic algorithm from the primer candidate obtained according to the process which makes a primer candidate the base sequence which acquires and processes the optimal base sequence for the differential displaying method from (a) manifestation gene database in the selection approach of a primer base sequence, and appear at the order of frequency, and the (b) aforementioned (a) process.
[Claim 2] It is the selection approach of the primer base sequence characterized by complementing by searching a homology with a genome base sequence and piling up with the same part when a manifestation gene database array does not fill 1000 bases with acquisition of said manifestation gene database base sequence in the selection approach of a primer base sequence according to claim 1.
[Claim 3] It is the selection approach of the primer base sequence characterized by performing the high-speed homology search which optimized decision of the frequency of said base sequence to full coincidence of a short array in the selection approach of a primer base sequence according to claim 1.
[Claim 4] In the selection equipment of a primer base sequence equipped with the processor which executes an instruction, the input device which inputs data, the store which memorizes the data inputted from this input device, and the output unit which outputs data (a) The input-process section which inputs the manifestation gene database read from said store, and a genome base sequence, (b) The new database creation processing section which processes data based on the data from this input-process section, and performs new database creation, (c) The processing section which performs each entry, complementary location count of a short base sequence, complementary location database creation, and a sort, (d) The primer candidate base sequence output section which outputs a primer candidate base sequence based on the data from this processing section, (e) The genetic algorithm processing section which performs genetic algorithm processing based on the data from this primer candidate base sequence output section, (f) Selection equipment of the primer base sequence characterized by providing the optimal primer candidate base sequence group output section which outputs an optimal primer candidate base sequence group based on the data from this genetic algorithm processing section.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the selection approach of the primer base sequence by the genetic algorithm which enables it to use in convenience the differential displaying method which is the experiment technique for the structure of a gene, a function or molecular evolution, and new gene acquisition, and its equipment.
[0002]
[Description of the Prior Art] The gene in a life is information which is known as DNA and by which the code was carried out to four kinds of bases on the organic chemistry matter. The gene base sequence on DNA is imprinted by RNA, and is translated into protein. A living thing leads a life, completing the body and coping with the changing external world, makes a descendant, and dies of the flow of the genetic information soon. A gene is that is, like engineering drawing of a life activity, or the guide of a life activity.
[0003] It has been determined in large quantities [the genome DNA sequence by which all genetic information is borne, and the actually discovered base sequence data of a gene], and quickly by the DNA sequencing technique in recent years.
[0004] A computer is introduced into molecular biology for management of these data, or analysis, and it is KOMPYUTEISHONARU. Molecular The fusion field of the information science and molecular biology which are called biotechnology ROJII (count molecular biology: computational molecular biology) or bioinfomatics (life information science: bioinformatics) is progressing.
[0005] Moreover, in the field of molecular biology, in order to search for gene expression, various approaches are developed until now and used. It is law. as one of them -- current -- being carried out widely -- PCR (polymerase chain reaction) -- the differential display (Differential Display) which is the kind -- law A primer with the base sequence considered to combine with more cDNA(s) at cDNA (group) created from mRNA, It is the laboratory procedure which performs the PCR method by the primer corresponding to the Poly-A array of a 3'-end, and they are coincidence and the outstanding technique for which it can search systematically about the manifestation about many genes.
[0006]
[Problem(s) to be Solved by the Invention] However, use of the database in the differential displaying method only presumed the class of gene from the database after the base sequence determination of the mainly obtained gene. Moreover, although there was a

program which selects the primer for the PCR methods from a single DNA sequence until now, the system which chooses the primer which covers more base sequences is difficult to realize using a current program, and it could not but depend for it on the rule of thumb until now about selection of the primer used by the differential displaying method. Moreover, when choosing which primer, there was little duplication and there were not a new gene's being chosen and a means to get to know.

[0007] In order that this invention may solve the above-mentioned trouble, may compare more generation products efficiently by the differential displaying method and may raise the probability of new gene acquisition, it aims at offering the selection approach of the optimal primer base sequence, and its equipment.

[0008]

[Means for Solving the Problem] In order to attain the above-mentioned purpose, this invention acquires and processes the optimal base sequence for the differential displaying method from a (a) manifestation gene database, and is made the process which selects the optimal primer group characterized by use of a genetic algorithm from the primer candidate obtained according to the process which makes a primer candidate the base sequence appearing [many] at the order of frequency, and the (b) aforementioned (a) process giving in the selection approach of a [1] primer base sequence.

[0009] [2] In the selection approach of the primer base sequence the above-mentioned [1] publication, when a manifestation gene database array does not fulfill 1000 bases, complement acquisition of said manifestation gene database base sequence by searching a homology with a genome base sequence and piling up with the same part.

[0010] [3] In the selection approach of the primer base sequence the above-mentioned [1] publication, decision of the frequency of said base sequence is made to perform the high-speed homology (similar) search optimized to full coincidence of a short array.

[0011] [4] In the selection equipment of a primer base sequence equipped with the processor which executes an instruction, the input device which inputs data, the store which memorizes the data inputted from this input device, and the output unit which outputs data The input-process section which inputs the manifestation gene database read from said store, and a genome base sequence, The new database creation processing section which processes data based on the data from this input-process section, and performs new database creation, The processing section which performs each entry, complementary location count of a short base sequence, complementary location database creation, and a sort, The primer candidate base sequence output section which outputs a primer candidate base sequence based on the data from this processing section, The genetic algorithm processing section which performs genetic algorithm processing based on the data from this primer candidate base sequence output section, The optimal primer candidate base sequence group output section which outputs an optimal primer candidate base sequence group based on the data from this genetic algorithm processing section is provided.

[0012] Since it constituted as mentioned above, the multiple selection of the optimal primer for the differential display which was difficult until now can be made.

[0013]

[Embodiment of the Invention] Hereafter, it explains to a detail, referring to a drawing about the gestalt of operation of this invention.

[0014] The selection structure-of-a-system Fig. of a primer base sequence in which

drawing 1 shows the example of this invention, and drawing 2 are the outline operation flow charts of the selection system of the primer base sequence.

[0015] As shown in drawing 1 , the selection system of the primer base sequence of this invention The processor (CPU/memory) 10 which executes an instruction, and the input unit 1 which inputs data, It has the store which memorizes the data inputted from this input device 1, and the output unit 2 which outputs data. That processor 10 Complementary location count of the input-process section 3 of an EST (manifestation gene) database and a genome base sequence, the new database creation processing section 4 by data processing based on an EST (manifestation gene) database genome base sequence, and an each entry and a short base sequence, It consists of complementary location database creation and the sorting application section 5, the primer candidate base sequence output section 6, the genetic algorithm processing section 7, and the optimal primer candidate base sequence group (put together) output section 8.

[0016] And the selection procedure of a primer base sequence performs the input of (1) EST (manifestation gene) database and a genome base sequence, as shown in drawing 2 (step S1). (2) New database creation processing by data processing based on the EST (manifestation gene) database and genome base sequence is performed (step S2). (3) Complementary location count of each entry and a short base sequence, complementary location database creation, and sorting application are performed (step S3). (4) Perform a primer candidate base sequence output (step S4), and (5) primer candidate base sequence data are acquired. Then, in the (6) step S5 which confirms whether make it termination (step S5), in YES, output primer candidate base sequence data, and it ends. in step S5, in NO, it progresses to the following step, genetic algorithm processing is performed (step S6), and (7) optimal primer candidate base sequence group (put together) is outputted to it -- it is made like (step S7).

[0017] The above-mentioned step S2 in drawing 2 - step S4 are shown in a detail, as shown in drawing 3 .

[0018] ** Set for an EST (manifestation gene) database and each entry, and write out the array for 3'-end 1000 base from what searches the keyword of a "3'-end" and agrees to another file. That is, by the differential displaying method, the array for 1000 bases is chosen from a 3'-end Poly-A array detectable with sufficient repeatability, and it writes out from all the cDNA base sequences of an EST (manifestation gene) database to another file (step S11).

[0019] ** Next, it is considered that precision falls, so that it becomes back from a base-sequence-determination start point before and behind 500 bases, even if a single EST array is long. Moreover, on the property of a present-day DNA sequencing technique, since Base G (guanine) is indicated to be X (decision impossible) in many cases, this is all amended with G (step S12).

[0020] ** Next, confirm whether there are some which are not filled into 1000 bases with the amended EST array a short thing and here (step S13).

[0021] ** In order that a short thing may measure and combine a complementary location from a genome database in the EST array subsequently amended in this way, perform a homology (similar) search (step S14).

[0022] In this case, as shown in drawing 4 , when the base which are not two or more complementation among 25 base sequences ahead of [of an EST array] base sequence determination appears, it repeats shifting the homology search starting position of a

genome and searching again. The optimization for finding the target heavy location can perform a homology search at a high speed very much compared with the usual search algorithm by the base only with these four kinds.
[0023] The sense of 5'- to 3'- is changed, and if required in the part except the intron under genome array (field omitted after an imprint), it will change for the array in a part for 1000 bases, and the above-mentioned file, and will write out from the location which carried out complementary coincidence. A part for the 1000 bases is considered as one entry. It is confirmed again whether, in that case, it is in agreement with an EST array. Moreover, about the EST array which does not carry out complementary coincidence, it does not carry out changing.
[0024] ** By this, as shown in drawing 8 , the new file for databases which specialized in the differential displaying method can be built (step S15).
[0025] ** next -- if it confirms whether the beginning to the new file for databases was completed and has ended about no entries -- step S11 -- return -- if it has ended, it will progress to the following step (step S16).
[0026] The following primer candidate base sequence retrieval is performed using this file for new databases.
[0027] A thing including more T arrays which do not include convenient conditions, i.e., palindrome prohibition, and T array which continued three or more in actually using it for an experiment by the differential displaying method chooses whether it is the number base (they are six bases or eight bases at a default) which carried out conditioning called exclusion, and is taken as the base sequence for performing primer matching.
[0028] ** Perform a homology (similar) search for how many times it comes out and a part for the sum total and what entry the inside of the short base sequence of the n-th power individual of 4 is among the above-mentioned file for new databases as a round robin.
[0029] It can refer to this homology search at a high speed very much rather than usual for trying to change the retrieval location of an entry, if one base is not the complementation, either, and the program which specialized in the short array. Moreover, whether one number base sequence carried out the complementation in which location under which entry saves further as a complementary database at another file (step S17).
[0030] ** complementary [among the number base sequences investigated for above-mentioned round robin] -- sort the bottom in an order from what has many entries (step S18).
[0031] ** Next, write out to a primer candidate file as a candidate of a primer base sequence (step S19). This can be seen directly and a primer can be chosen.
[0032] (10) next -- if it checks and writes whether the beginning processing to a primer candidate file was completed and appearance is not completed -- step S17 -- return -- if it wrote and appearance is completed -- and -- ** -- carry out (step S20).
[0033] When it is the experiment which does not need to select the following primer groups, it is also possible to choose a primer base sequence from this file suitably.
[0034] Next, step S5 and step S6 which were shown in drawing 2 are explained to a detail. That is, the combination of the primer which has more entry gene expression confirmed by the genetic algorithm using the above-mentioned primer candidate file and a complementary database is chosen. When this combination is that [number (base) ****** (step S16 reference)] of 4 (base), and combines as a round robin and the original primer candidate array makes it calculate, astronomical count power is needed. By using a genetic

algorithm for this, the speed which chooses the optimal primer group increases by leaps and bounds, and becomes computable also by the comparatively small-scale computer.
[0035] Drawing 5 and drawing 6 are the explanatory views about the output method of the combination of the primer which can confirm much entry gene expression from genetic algorithm processing.
[0036] as a concrete algorithm -- (1) -- first, it entry-evaluation-array-creates and initializes (step S21).
[0037] At this time, as shown in drawing 7 , the number array of entries which first set to 1 1 bit corresponding to the entry number which made 1 bit every primer candidate base sequence one entry, and carried out the complementation, and set except [its] to 0 is made to the above-mentioned complementation database creation time.
[0038] This is for compressing a storage region and there is no need of adhering to making one entry into 1 bit. The direction calculated by 1 byte can accelerate depending on an algorithm.
[0039] Moreover, the combination of a primer candidate base sequence some (a default 8), i.e., the row of a two-dimensional array, is chosen from dozens 100 (a default is 100) or more with the random number. The thing which had the same array in the same combination (row), and the thing of the same combination except. (They are dozens about "combination", such as "AGAAAA CTGGAA CATTAA") .
[0040] (2) Next, the two-dimensional array for genetic algorithms (refer to drawing 7) calculates to what entry the complementation is carried out about each line (step S22). That is, the number of evaluations of to what entry to have carried out the complementation is calculated. An OR operation is applied to the number array of entries created above about each primer candidate base sequence in combination, and a bit counts the number of 1. That is, a primer candidate base sequence group calculates under how many entries it is concretely in agreement.
[0041] Since it is an OR operation, it is high-speed, and since it shifts also about the number array of entries and a number is only counted, it can calculate at a high speed. In a genetic algorithm, since the part which spends time amount most calculates the number of evaluations, the meaning which accelerates this is large.
[0042] (3) Next, rearrange each line from a number (the number of evaluations) of high orders which carried out the complementation (step S23).
[0043] (4) Next, perform the recombination operation of a genetic algorithm. A primer candidate group is first rearranged from the high order of the number of evaluations. Since this uses a hash, it is high-speed. About the number high orders 1/4 of evaluations of a primer group, it places as it is. Moreover, it attaches by 1/2 from the high orders 1/4 of the number of evaluations of two-dimensional array, and one of each primer candidate array of the combination which it was random and was chosen from high orders 1/4, and the combination chosen from another high orders 1/4 is chosen by random numbers. The combination of low order 1/2 chooses a base sequence candidate by random numbers completely. Completely, a thing with the same primer candidate array and the same combination are excepted, and are newly chosen by random numbers (step S24). In addition, the primer candidate array file is shown in drawing 8 .
[0044] (5) Next, repeat from step S22 to the step S24 hundreds times or more (a default is 200 times) (step S25). Thereby, the optimal primer candidate base sequence group can be selected.

[0045] (6) Next, if step S25 is satisfied, the optimal primer candidate base sequence group will be outputted in order of a high order (step S26). In addition, the optimal primer candidate array group file is shown in drawing 9 .
[0046] In addition, this invention is not limited to the above-mentioned example, and based on the meaning of this invention, many deformation is possible for it and it does not eliminate these from the range of this invention.
[0047]
[Effect of the Invention] As mentioned above, according to this invention, the following effectiveness can be done so as explained to the detail.
[0048] (A) The concrete base sequence considered to be the the best for a primer can be outputted sequentially from a high order. Although it is thought by different species that the base sequences of the 3'-end of a gene cDNA base sequence differ delicately, the optimal primer can be obtained by comparatively short time amount by calculating in the kind of cDNA database.
[0049] (B) If the differential displaying method is performed using the primer group chosen by the genetic algorithm, it is several [only] primers and a theory top is possible for investigating the existence of a manifestation of almost all cDNA(s). It is actually Yeast Saccharomices. When the manifestation gene database of Cereviciae was used and the program of this invention was performed, it turned out also by the single base sequence that there are many entries and coincidence parts. But and it is predicted among all the EST (manifestation gene) database of yeast that it combines with 70% or more of the entry the 3'-end was indicated to be. Generally it is thought about ten% rather than the number of entries which is in agreement by the primer of a commercial rule of thumb that there is much this.
[0050] (C) If the combination of the best primer group obtained by this system is used, in the case of yeast, it will be thought in the combination of eight kinds of primers of six base sequences that 92% or more of manifestation of the entry in [all] EST (manifestation gene) can be investigated. This is very cheap and simpler than the approach of performing PCR using the primer of the conventional immense amount (thousands), and investigating by the DNA array.
[0051] Current and the genome project of various kinds are advancing. Since this system can be performed to the database created newly in the future, it should become the functional analysis of a gene and an aid of new gene acquisition by the differential displaying method.

[Translation done.]

* NOTICES *

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the selection structure-of-a-system Fig. of a primer base sequence showing the example of this invention.
[Drawing 2] It is the selection flow chart of the outline primer base sequence of this invention.
[Drawing 3] It is the selection flow chart of the primer base sequence which shows the example of this invention.
[Drawing 4] It is the explanatory view of the homology search in selection of the primer base sequence which shows the example of this invention.
[Drawing 5] It is the flow chart of the genetic algorithm in selection of the primer base sequence which shows the example of this invention.
[Drawing 6] It is the explanatory view of the genetic algorithm in selection of the primer base sequence which shows the example of this invention.
[Drawing 7] It is the explanatory view of the genetic algorithm processed-data structure in selection of the primer base sequence which shows the example of this invention.
[Drawing 8] It is drawing showing the example of a primer candidate file in selection of the primer base sequence which shows the example of this invention.
[Drawing 9] It is drawing showing the example of an optimal primer candidate base sequence group file in selection of the primer base sequence which shows the example of this invention.
[Description of Notations]
1 Input Unit
2 Output Unit
3 Input-Process Section of EST (Manifestation Gene) Database and Genome Base Sequence
4 New Database Creation Processing Section
5 Complementary Location Count of Each Entry and Short Base Sequence, Complementary Location Database Creation, and Sorting Application Section
6 Primer Candidate Base Sequence Output Section
7 Genetic Algorithm Processing Section
8 Optimal Primer Candidate Base Sequence Group (Put Together) Output Section
10 Processor (CPU/Memory)

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2000－308487
（P2000－308487A）
(43)公開日　平成12年11月７日(2000.11.7)

(51)Int.Cl.'		識別記号	FI		テーマコード'(参考)
Ｃ１２Ｎ	15/09	ＺＮＡ	Ｃ１２Ｎ	15/00	ＺＮＡＡ　４Ｂ０２４
Ｃ１２Ｑ	1/68		Ｃ１２Ｑ	1/68	Ａ　４Ｂ０６３
Ｇ０６Ｆ	15/18	５５０	Ｇ０６Ｆ	15/18	５５０Ｃ　５Ｂ０７５
	17/30			15/40	３７０Ｆ

審査請求　未請求　請求項の数４　ＯＬ　（全 13 頁）

(21)出願番号　特願平11－88410

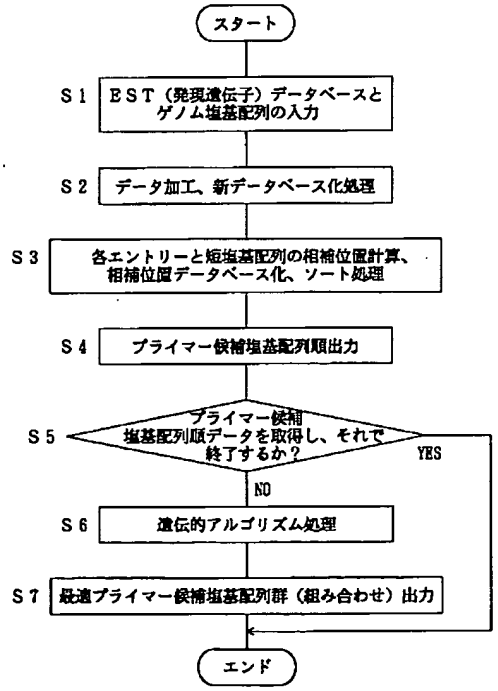
(22)出願日　平成11年３月30日(1999.3.30)

(71)出願人　396020800
科学技術振興事業団
埼玉県川口市本町４丁目１番８号
(72)発明者　村上　大勇
埼玉県草加市谷塚町49－３　ライオンズガーデン谷塚501
(72)発明者　金澤　一郎
東京都大田区田園調布１－47－17
(72)発明者　鈴木　高史
愛知県名古屋市瑞穂区瑞穂町宇川澄１
(74)代理人　100089635
弁理士　清水　守

最終頁に続く

(54)【発明の名称】　プライマー塩基配列の選定方法及びその装置

(57)【要約】
【課題】　ディファレンシャル・ディスプレイ法で生成産物をより多く効率的に比較し、新規遺伝子取得の確率を高めるため、最適なプライマー塩基配列の選定方法及びその装置を提供する。
【解決手段】　プライマー塩基配列の選定方法において、（ａ）発現遺伝子データベースからディファレンシャル・ディスプレイ法に最適な塩基配列を取得して加工し、多く現れる塩基配列を頻度順にプライマー候補とする工程と、（ｂ）前記（ａ）工程により得られたプライマー候補から遺伝的アルゴリズムの使用を特徴とする最適プライマー群を選定する工程とを施すようにしたものである。



【特許請求の範囲】
【請求項１】　プライマー塩基配列の選定方法において、（ａ）発現遺伝子データベースからディファレンシャル・ディスプレイ法に最適な塩基配列を取得して加工し、多く現れる塩基配列を頻度順にプライマー候補とする工程と、（ｂ）前記（ａ）工程により得られたプライマー候補から遺伝的アルゴリズムの使用を特徴とする最適プライマー群を選定する工程とを施すことを特徴とするプライマー塩基配列の選定方法。
【請求項２】　請求項１記載のプライマー塩基配列の選定方法において、前記発現遺伝子データベース塩基配列の取得は、発現遺伝子データベース配列が１０００塩基に満たない場合にゲノム塩基配列とのホモロジーを検索して同一部分と重ね合わせることにより補完することを特徴とするプライマー塩基配列の選定方法。
【請求項３】　請求項１記載のプライマー塩基配列の選定方法において、前記塩基配列の頻度の確定は、短い配列の完全一致用にオプティマイズした高速ホモロジーサーチを行うことを特徴とするプライマー塩基配列の選定方法。
【請求項４】　命令を実行するプロセッサと、データを入力する入力装置と、該入力装置から入力されたデータを記憶する記憶装置と、データを出力する出力装置とを備えるプライマー塩基配列の選定装置において、（ａ）前記記憶装置から読み出される発現遺伝子データベースとゲノム塩基配列とを入力する入力処理部と、（ｂ）該入力処理部からのデータに基づいてデータの加工を行い新しいデータベース化を行う新データベース化処理部と、（ｃ）各エントリーと短塩基配列の相補位置計算と相補位置データベース化及びソートを行う処理部と、（ｄ）該処理部からのデータに基づいてプライマー候補塩基配列順を出力するプライマー候補塩基配列順出力部と、（ｅ）該プライマー候補塩基配列順出力部からのデータに基づいて遺伝的アルゴリズム処理を行う遺伝的アルゴリズム処理部と、（ｆ）該遺伝的アルゴリズム処理部からのデータに基づいて最適プライマー候補塩基配列群を出力する最適プライマー候補塩基配列群出力部とを具備することを特徴とするプライマー塩基配列の選定装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、遺伝子の構造や機能、または分子進化、新規遺伝子取得のための実験手法であるディファレンシャル・ディスプレイ法を、より利便的に使用できるようにする、遺伝的アルゴリズムによるプライマー塩基配列の選定方法及びその装置に関するものである。
【０００２】
【従来の技術】生命における遺伝子は、ＤＮＡとして知られる、有機化学物質上の４種類の塩基にコードされた情報である。ＤＮＡ上の遺伝子塩基配列はＲＮＡに転写され、たんぱく質に翻訳される。その遺伝情報の流れにより、生物は体を作りあげ、変化する外界に対処しながら生活を営み、子孫を作りやがて死に至る。つまり遺伝子は生命活動の設計図、または生命活動の手引書のようなものである。
【０００３】近年のＤＮＡ塩基配列決定技術によって、一切の遺伝情報が担われているゲノムＤＮＡ塩基配列、また実際に発現している遺伝子の塩基配列データが大量、急速に決定されてきている。
【０００４】これらのデータの管理や解析のために、コンピュータが分子生物学に導入され、コンピュテイショナル　モレキュラー　バイオロジィー（計算分子生物学：ｃｏｍｐｕｔａｔｉｏｎａｌ　ｍｏｌｅｃｕｌａｒ　ｂｉｏｌｏｇｙ）、あるいはバイオインフォマティックス（生命情報科学：ｂｉｏｉｎｆｏｒｍａｔｉｃｓ）と呼ばれている、情報科学と分子生物学との融合領域が発達してきている。
【０００５】また、分子生物学の分野で、遺伝子発現を探索するために様々な方法がこれまで開発され使われている。その一つとして現在広く行われているのがＰＣＲ（ポリメラーゼ・チェイン・リアクション）法であり、その一種であるディファレンシャル・ディスプレイ（Ｄｉｆｆｅｒｅｎｔｉａｌ　Ｄｉｓｐｌａｙ）法は、ｍＲＮＡより作成したｃＤＮＡ（群）に、より多くのｃＤＮＡと結合すると思われる塩基配列を持ったプライマーと、3’－末端のＰｏｌｙ－Ａ配列に対応するプライマーとでＰＣＲ法を行う実験法であり、多数の遺伝子について、その発現を同時かつ系統的に探索できる優れた手法である。
【０００６】
【発明が解決しようとする課題】しかしながら、ディファレンシャル・ディスプレイ法におけるデータベースの使用は、主に得られた遺伝子の塩基配列決定後にデータベースから遺伝子の種類を推定するだけであった。また、これまで単一のＤＮＡ塩基配列よりＰＣＲ法用のプライマーを選定するプログラムはあったが、より多くの塩基配列をカバーするプライマーを選択するシステムは、現在のプログラムを使って実現するのが困難であり、ディファレンシャル・ディスプレイ法で使用するプライマーの選択については、これまでは経験則に頼るしかなかった。また、どのプライマーを選べば重複が少なく、新規遺伝子を選べるのかも、知る手段が無かった。
【０００７】本発明は、上記の問題点を解決し、ディファレンシャル・ディスプレイ法で生成産物をより多く効率的に比較し、新規遺伝子取得の確率を高めるため、最適なプライマー塩基配列の選定方法及びその装置を提供することを目的とする。
【０００８】
【課題を解決するための手段】本発明は、上記目的を達

成するために、
〔１〕プライマー塩基配列の選定方法において、（ａ）発現遺伝子データベースからディファレンシャル・ディスプレイ法に最適な塩基配列を取得して加工し、多く現れる塩基配列を頻度順にプライマー候補とする工程と、（ｂ）前記（ａ）工程により得られたプライマー候補から遺伝的アルゴリズムの使用を特徴とする最適プライマー群を選定する工程とを施すようにしたものである。
【０００９】〔２〕上記〔１〕記載のプライマー塩基配列の選定方法において、前記発現遺伝子データベース塩基配列の取得は、発現遺伝子データベース配列が１０００塩基に満たない場合にゲノム塩基配列とのホモロジーを検索して同一部分と重ね合わせることにより補完するようにしたものである。
【００１０】〔３〕上記〔１〕記載のプライマー塩基配列の選定方法において、前記塩基配列の頻度の確定は、短い配列の完全一致用にオプティマイズした高速ホモロジー（類似）サーチを行うようにしたものである。
【００１１】〔４〕命令を実行するプロセッサと、データを入力する入力装置と、この入力装置から入力されたデータを記憶する記憶装置と、データを出力する出力装置とを備えるプライマー塩基配列の選定装置において、前記記憶装置から読み出される発現遺伝子データベースとゲノム塩基配列とを入力する入力処理部と、この入力処理部からのデータに基づいてデータの加工を行い新しいデータベース化を行う新データベース化処理部と、各エントリーと短塩基配列の相補位置計算と相補位置データベース化及びソートを行う処理部と、この処理部からのデータに基づいてプライマー候補塩基配列順を出力するプライマー候補塩基配列順出力部と、このプライマー候補塩基配列順出力部からのデータに基づいて遺伝的アルゴリズム処理を行う遺伝的アルゴリズム処理部と、この遺伝的アルゴリズム処理部からのデータに基づいて最適プライマー候補塩基配列群を出力する最適プライマー候補塩基配列群出力部とを具備するようにしたものである。
【００１２】上記のように構成したので、これまで困難であったディファレンシャル・ディスプレイに最適なプライマーを複数選択することができる。
【００１３】
【発明の実施の形態】以下、本発明の実施の形態について図面を参照しながら詳細に説明する。
【００１４】図１は本発明の実施例を示すプライマー塩基配列の選定システムの構成図、図２はそのプライマー塩基配列の選定システムの概略動作フローチャートである。
【００１５】図１に示すように、本発明のプライマー塩基配列の選定システムは、命令を実行する処理装置（ＣＰＵ／メモリ）１０と、データを入力する入力装置１と、この入力装置１から入力されたデータを記憶する記憶装置と、データを出力する出力装置２とを備え、その処理装置１０は、ＥＳＴ（発現遺伝子）データベースとゲノム塩基配列の入力処理部３、ＥＳＴ（発現遺伝子）データベース・ゲノム塩基配列に基づいたデータ加工による新データベース化処理部４、各エントリーと短塩基配列の相補位置計算、相補位置データベース化及びソート処理部５、プライマー候補塩基配列順出力部６、遺伝的アルゴリズム処理部７、最適プライマー候補塩基配列群（組み合わせ）出力部８からなる。
【００１６】そして、プライマー塩基配列の選定手順は、図２に示すように、（１）ＥＳＴ（発現遺伝子）データベースとゲノム塩基配列の入力を行い（ステップＳ１）、（２）そのＥＳＴ（発現遺伝子）データベースとゲノム塩基配列に基づいたデータ加工による新データベース化処理を行い（ステップＳ２）、（３）各エントリーと短塩基配列の相補位置計算、相補位置データベース化及びソート処理を行い（ステップＳ３）、（４）プライマー候補塩基配列順出力を行い（ステップＳ４）、（５）プライマー候補塩基配列順データを取得し、それで終了にするか否かをチェックする（ステップＳ５）、（６）ステップＳ５においてＹＥＳの場合にはプライマー候補塩基配列順データを出力して終了し、ステップＳ５においてＮＯの場合には、次のステップへ進み、遺伝的アルゴリズム処理を行い（ステップＳ６）、（７）最適プライマー候補塩基配列群（組み合わせ）を出力する（ステップＳ７）ようにしている。
【００１７】図２における上記ステップＳ２～ステップＳ４は、詳細には、図３に示すように、示される。
【００１８】①ＥＳＴ（発現遺伝子）データベース、各エントリーにおいて、「３′－末端」のキーワードを検索、合致するものから、３′－末端１０００塩基分の配列を別ファイルに書き出す。つまり、ＥＳＴ（発現遺伝子）データベースの、すべてのｃＤＮＡ塩基配列から、ディファレンシャル・ディスプレイ法によって再現性良く検出できる３′－末端Ｐｏｌｙ－Ａ配列より１０００塩基分の配列を選択し、別ファイルに書き出す（ステップＳ１１）。
【００１９】②次に、単一のＥＳＴ配列は、長くても５００塩基前後で、塩基配列決定開始点から後方になるほど精度が落ちると考えられる。また、現代のＤＮＡ塩基配列決定技術の性質上、塩基Ｇ（グアニン）がＸ（決定不能）と記載されることが多いので、これをすべてＧと補正する（ステップＳ１２）。
【００２０】③次に、補正したＥＳＴ配列で短いもの、ここでは、１０００塩基に満たないものがあるか否かをチェックする（ステップＳ１３）。
【００２１】④次いで、このように補正したＥＳＴ配列で短いものは、ゲノムデータベースから、相補位置を測定して組み合わせるため、ホモロジー（類似）サーチを行う（ステップＳ１４）。

【００２２】この場合、図４に示すように、ＥＳＴ配列の塩基配列決定前方の２５塩基配列中２箇所以上の相補でない塩基が現れた場合、ゲノムのホモロジーサーチ開始位置をずらして再度サーチを行うことを繰り返す。この４種類しかない塩基で目的の重ね位置を見つけるためのオプティマイズにより、通常のサーチアルゴリズムに比べて非常に高速にホモロジーサーチを行うことができる。

【００２３】相補一致した位置から、ゲノム配列中のイントロン（転写後に切り捨てられる領域）を除く部分を、必要ならば、5′－から3′－の向きを変換して、１０００塩基分、上記ファイル中の配列と入れ替えて書き出す。その１０００塩基分を１エントリーとする。その際に、ＥＳＴ配列と一致しているか再度チェックする。また相補一致しないＥＳＴ配列については、入れ替えることをしない。

【００２４】⑤これによって、例えば、図８に示すように、ディファレンシャル・ディスプレイ法に特化した、新たなデータベース用ファイルを構築できる（ステップＳ１５）。

【００２５】⑥次に、全てのエントリーについて新たなデータベース用ファイルへの書き出しが終了したか否かをチェックし、終了していなければ、ステップＳ１１へ戻り、終了していれば、次のステップへ進む（ステップＳ１６）。

【００２６】以下のプライマー候補塩基配列探索は、この新データベース用ファイルを使用して行う。

【００２７】ディファレンシャル・ディスプレイ法で実験に実際に使用するのに都合の良い条件、つまり、パリンドローム禁止、３つ以上続いたＴ配列を含まない、より多いＴ配列を含んだものは除外、という条件付けをした数塩基（デフォルトで６塩基または８塩基）であるか否かを選択し、プライマーマッチングを行うための塩基配列とする。

【００２８】⑦４のｎ乗個の短塩基配列の中が、上記新データベース用ファイル中、何回出てくるのかその合計と何エントリー分であるかを、総当りでホモロジー（類似）サーチを行う。

【００２９】このホモロジーサーチでは、１塩基でも相補でないとエントリーの検索位置を換えるようにしているのと、短い配列用に特化したプログラムのため、通常よりも非常に高速に検索できる。また、一つの数塩基配列が、どのエントリー中の、どの位置で相補したかは、相補データベースとして、さらに別ファイルに保存しておく（ステップＳ１７）。

【００３０】⑧上記総当たりで調べた数塩基配列のうち、相補したエントリーが多いものから順番にソートする（ステップＳ１８）。

【００３１】⑨次に、プライマー塩基配列の候補としてプライマー候補ファイルに書き出す（ステップＳ１９）。これを直接見てプライマーを選択することができる。

【００３２】(10)次に、プライマー候補ファイルへの書き出し処理が終了したか否かをチェックして、書き出が終了していなかったらステップＳ１７へ戻り、書き出が終了していたらエンドとする（ステップＳ２０）。

【００３３】以下のプライマー群を選定する必要がない実験の場合、このファイルからプライマー塩基配列を適当に選択することも可能である。

【００３４】次に、図２に示したステップＳ５およびステップＳ６について詳細に説明する。すなわち、遺伝的アルゴリズムによって、上記プライマー候補ファイルと相補データベースを利用して、より多くのエントリー遺伝子の発現を確かめられるプライマーの組み合わせを選ぶ。この組み合わせは、元のプライマー候補配列が４（塩基）の数（塩基）乗ある（ステップＳ１６参照）ので、総当たりで組み合わせ計算させると天文学的な計算パワーが必要になる。これを遺伝的アルゴリズムを利用することにより、最適プライマー群を選択するスピードが飛躍的に上がり、比較的小規模の計算機でも計算可能になる。

【００３５】図５及び図６は遺伝的アルゴリズム処理とより多くのエントリー遺伝子の発現を確かめられるプライマーの組み合わせの出力方法についての説明図である。

【００３６】具体的なアルゴリズムとして、

（１）まず、エントリー評価配列作成・初期化する（ステップＳ２１）。

【００３７】この時、図７に示すように、まず、各プライマー候補塩基配列ごと、１つのエントリーを１ビットとし相補したエントリー番号に対応する１ビットを１にし、それ以外を０にしたエントリー数配列を上記相補データベース作成時に作っておく。

【００３８】これは記憶領域を圧縮するためで、１つのエントリーを１ビットとすることにこだわる必要は無い。１バイトで計算した方が、アルゴリズムによっては高速化できる。

【００３９】また、乱数により、プライマー候補塩基配列数個（デフォルトでは８）の組み合わせ、つまり二次元配列の横列を数十から百以上（デフォルトは１００）選んでおく。同じ組み合わせ（横列）で、同一の配列を持ったもの、また、同一組み合わせのものは除外する。（“ＡＧＡＡＡＡ　ＣＴＧＧＡＡ　ＣＡＴＴＡＡ”等の“組み合わせ”を数十）。

【００４０】（２）次に、遺伝的アルゴリズム用２次元配列（図７参照）が、各行について何エントリーと相補するのか計算を行う（ステップＳ２２）。つまり、何エントリーと相補したかの評価数を計算する。組み合わせ中の各プライマー候補塩基配列について上記で作成したエントリー数配列にＯＲ演算をかけ、ビットが１の数を

数える。つまり、プライマー候補塩基配列群が、具体的にいくつのエントリー中と一致するかを計算する。

【００４１】ＯＲ演算なので高速であり、エントリー数配列についてもシフトして数を数えるだけなので高速に演算できる。遺伝的アルゴリズムでは、評価数を計算するのが一番時間を費やす部分なので、ここを高速化する意義は大きい。

【００４２】（３）次に、相補した数（評価数）の上位から、各行を並べ替える（ステップＳ２３）。

【００４３】（４）次に、遺伝的アルゴリズムの組換え演算を行う。まず評価数の上位からプライマー候補群を並べ換える。これはハッシュを使うので、高速である。プライマー群の評価数上位１／４については、そのまま置く。また２次元配列の評価数の上位１／４から１／２までについては、上位１／４からランダムで選んだ組み合わせと、別の上位１／４から選んだ組み合わせの各プライマー候補配列のどちらかを乱数で選ぶ。下位１／２の組み合わせは、完全に乱数で塩基配列候補を選ぶ。同一のプライマー候補配列を持ったもの、また、完全に同一の組み合わせは除外して、新たに乱数で選ぶ（ステップＳ２４）。なお、図８にプライマー候補配列ファイルが示されている。

【００４４】（５）次に、ステップＳ２２からステップＳ２４までを数百回以上（デフォルトは２００回）繰り返す（ステップＳ２５）。これにより、最適なプライマー候補塩基配列群を選定することができる。

【００４５】（６）次に、ステップＳ２５を満足したら、最適なプライマー候補塩基配列群を上位順に出力する（ステップＳ２６）。なお、図９に最適プライマー候補配列群ファイルが示されている。

【００４６】なお、本発明は上記実施例に限定されるものではなく、本発明の趣旨に基づいて数々の変形が可能であり、これらを本発明の範囲から排除するものではない。

【００４７】

【発明の効果】以上、詳細に説明したように、本発明によれば、以下のような効果を奏することができる。

【００４８】（Ａ）プライマーに最適と思われる具体的な塩基配列を上位から順に出力することができる。異なる種では遺伝子ｃＤＮＡ塩基配列の３’ー末端の塩基配列は微妙に異なると考えられるが、その種のｃＤＮＡデータベースで計算を行うことで、最適なプライマーを比較的短い時間で得ることができる。

【００４９】（Ｂ）遺伝的アルゴリズムで選ばれたプライマー群を利用してディファレンシャル・ディスプレイ法を行えば、わずか数個のプライマーで、ほとんどすべてのｃＤＮＡの発現の有無を調べることが、理論上可能である。実際に、酵母Ｓａｃｃｈａｒｏｍｉｃｅｓ Ｃｅｒｅｖｉｃｉａｅの発現遺伝子データベースを使用して、本発明のプログラムを実行したところ、単一の塩基配列でも、多くのエントリーと一致個所があることが分かった。もっとも多いもので、酵母の全ＥＳＴ（発現遺伝子）データベース中、３’ー末端が記載されたエントリーの７０％以上と結合すると予測される。これは市販の経験則のプライマーで一致するエントリー数よりも、一般的に１０数％は多いと考えられる。

【００５０】（Ｃ）本システムで得られた最上のプライマー群の組み合わせを利用すると、酵母の場合、６塩基配列のプライマー８種類の組み合わせで、全ＥＳＴ（発現遺伝子）中のエントリーの９２％以上の発現を調べられると考えられる。これは、従来の莫大な量（数千）のプライマーを用いてＰＣＲを行い、ＤＮＡアレーで調べる方法よりも非常に安価で簡便である。

【００５１】現在、さまざまな種のゲノムプロジェクトが進行中である。本システムは将来的に、新規に作成されたデータベースに対して実行できるので、ディファレンシャル・ディスプレイ法による遺伝子の機能解析と新規遺伝子取得の一助になるはずである。

【図面の簡単な説明】

【図１】本発明の実施例を示すプライマー塩基配列の選定システムの構成図である。

【図２】本発明の概略プライマー塩基配列の選定フローチャートである。

【図３】本発明の実施例を示すプライマー塩基配列の選定フローチャートである。

【図４】本発明の実施例を示すプライマー塩基配列の選定におけるホモロジーサーチの説明図である。

【図５】本発明の実施例を示すプライマー塩基配列の選定における遺伝的アルゴリズムのフローチャートである。

【図６】本発明の実施例を示すプライマー塩基配列の選定における遺伝的アルゴリズムの説明図である。

【図７】本発明の実施例を示すプライマー塩基配列の選定における遺伝的アルゴリズム処理データ構造の説明図である。

【図８】本発明の実施例を示すプライマー塩基配列の選定におけるプライマー候補ファイル例を示す図である。

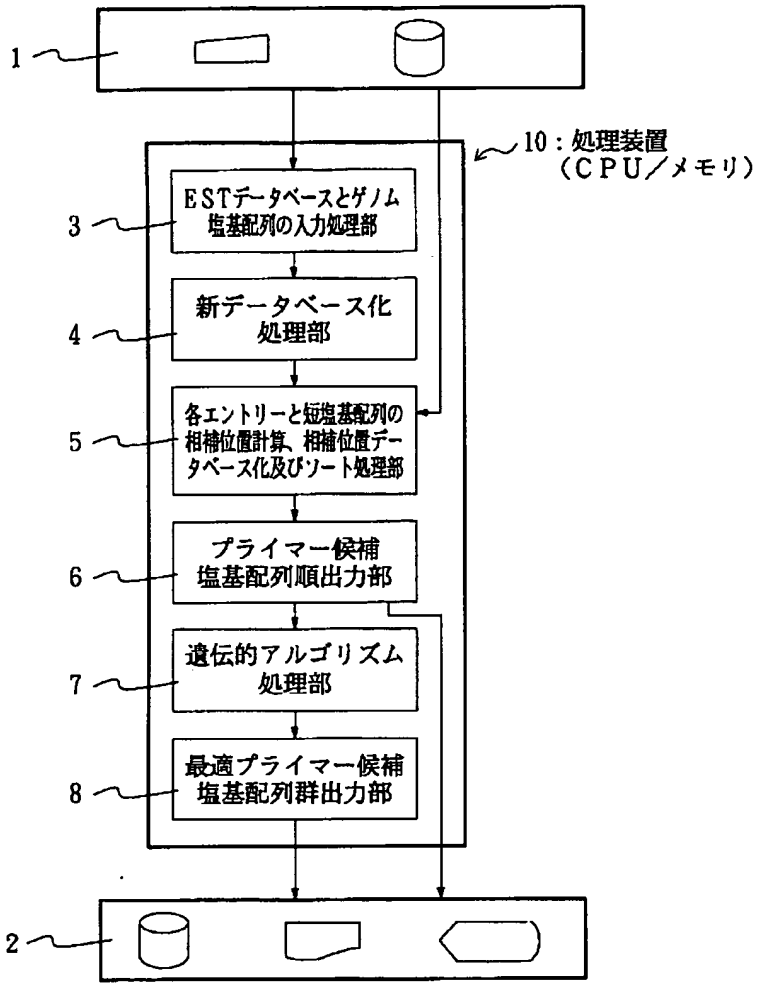
【図９】本発明の実施例を示すプライマー塩基配列の選定における最適プライマー候補塩基配列群ファイル例を示す図である。

【符号の説明】

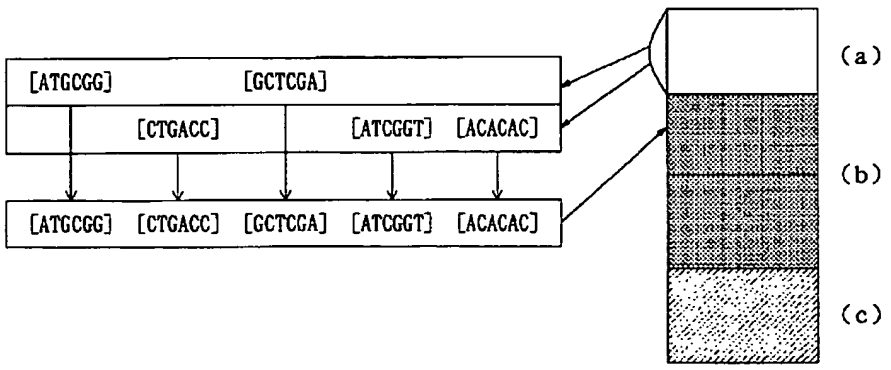
１　入力装置
２　出力装置
３　ＥＳＴ（発現遺伝子）データベースとゲノム塩基配列の入力処理部
４　新データベース化処理部
５　各エントリーと短塩基配列の相補位置計算、相補位置データベース化及びソート処理部
６　プライマー候補塩基配列順出力部
７　遺伝的アルゴリズム処理部

８　最適プライマー候補塩基配列群（組み合わせ）出力部　　１０　処理装置（ＣＰＵ／メモリ）

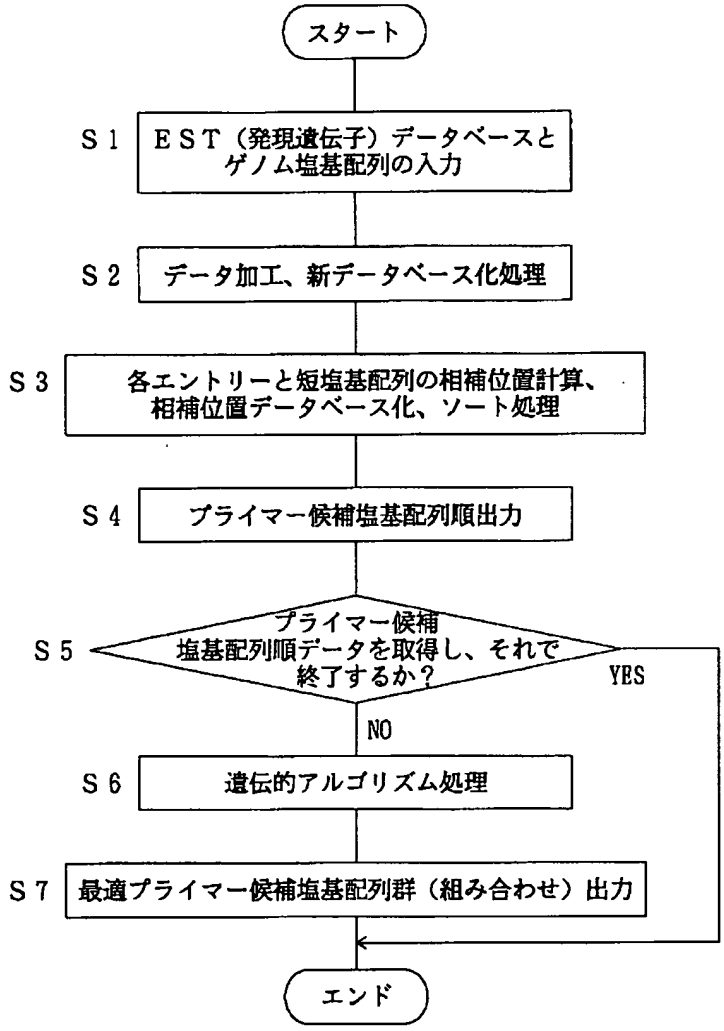
【図１】



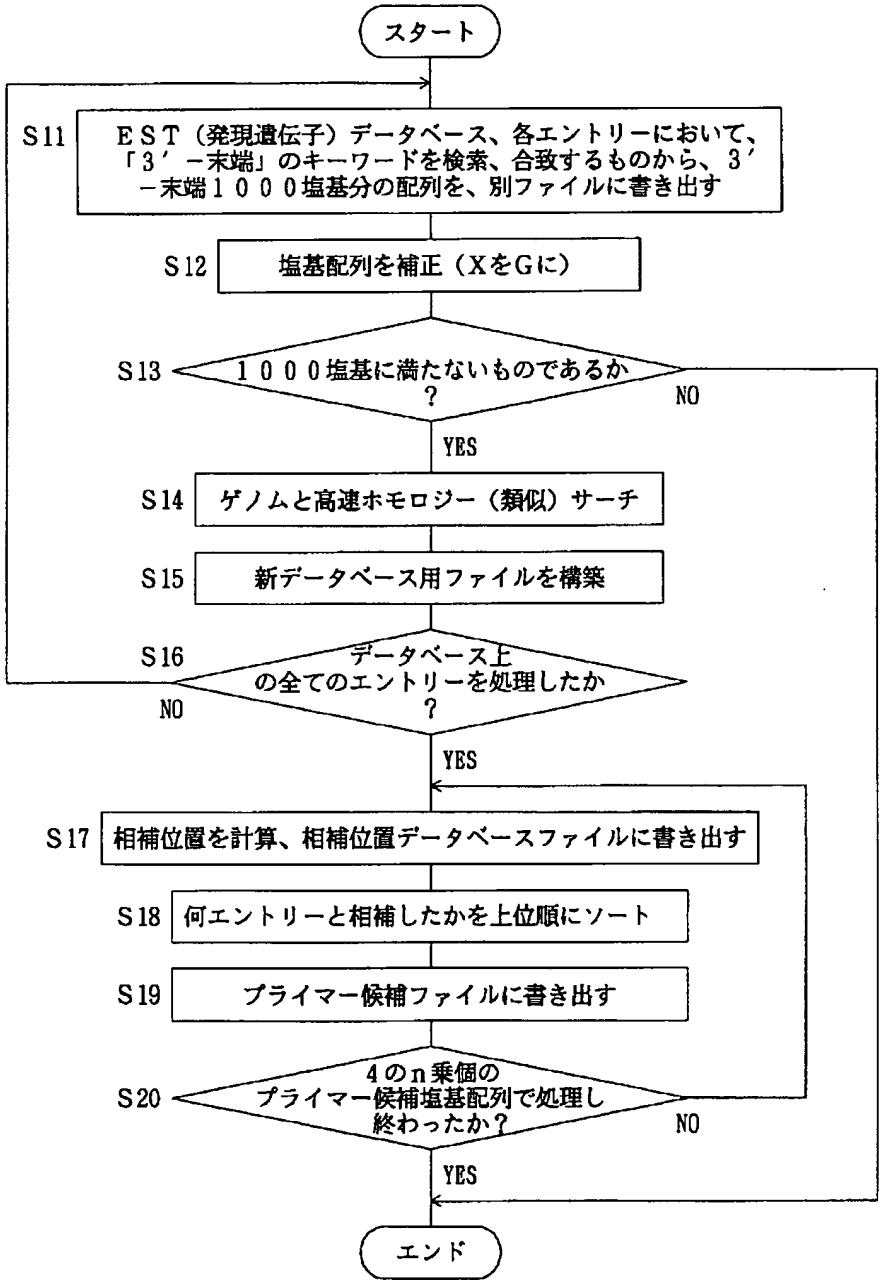
【図６】



【図２】


スタート
S１
ＥＳＴ（発現遺伝子）データベースと
ゲノム塩基配列の入力
S２
データ加工、新データベース化処理
S３
各エントリーと短塩基配列の相補位置計算、
相補位置データベース化、ソート処理
S４
プライマー候補塩基配列順出力
S５
プライマー候補
塩基配列順データを取得し、それで
終了するか？
YES
NO
S６
遺伝的アルゴリズム処理
S７
最適プライマー候補塩基配列群（組み合わせ）出力
エンド

【図３】

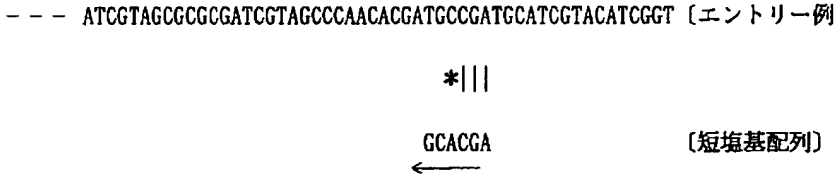

スタート
S11
ＥＳＴ（発現遺伝子）データベース、各エントリーにおいて、「３’－末端」のキーワードを検索、合致するものから、３’－末端１０００塩基分の配列を、別ファイルに書き出す
S12
塩基配列を補正（ＸをＧに）
S13
１０００塩基に満たないものであるか？
NO
YES
S14
ゲノムと高速ホモロジー（類似）サーチ
S15
新データベース用ファイルを構築
S16
データベース上の全てのエントリーを処理したか？
NO
YES
S17
相補位置を計算、相補位置データベースファイルに書き出す
S18
何エントリーと相補したかを上位順にソート
S19
プライマー候補ファイルに書き出す
S20
４のｎ乗個のプライマー候補塩基配列で処理し終わったか？
NO
YES
エンド

【図４】

（ａ）ゲノム塩基配列とＥＳＴエントリーとのホモロジーサーチ

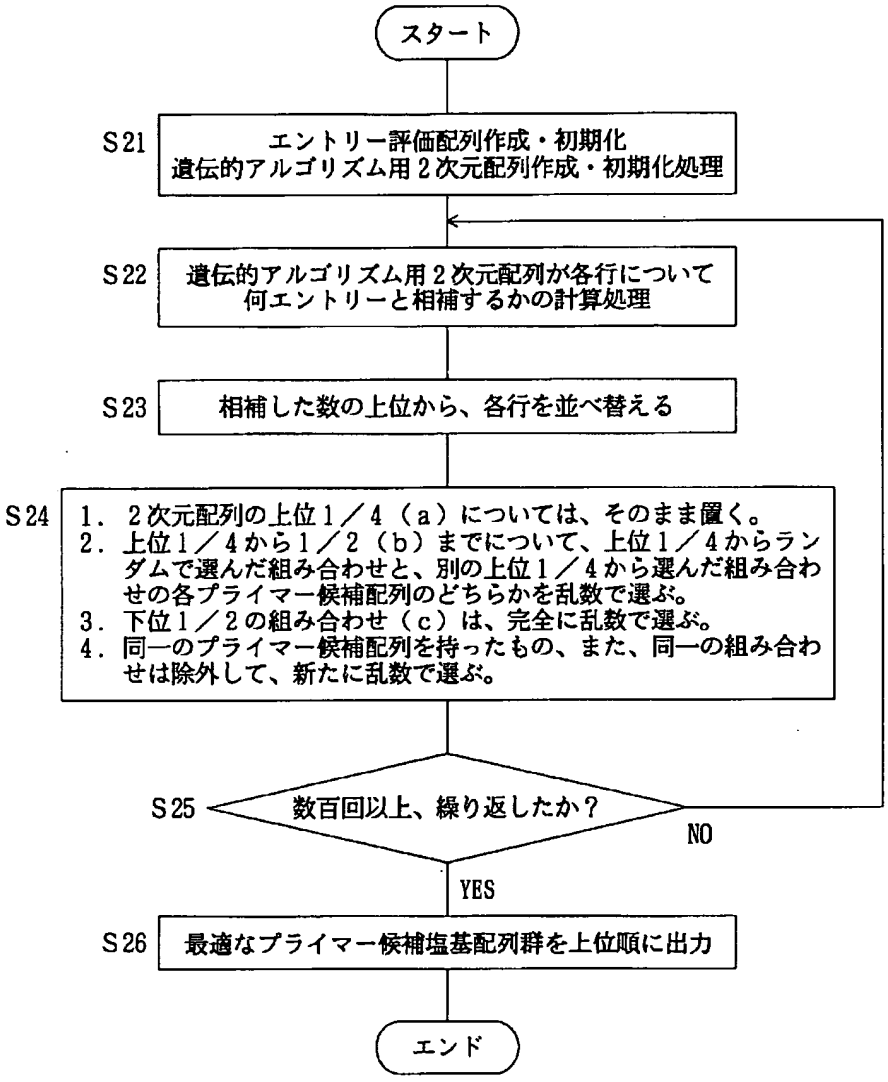
・２個ミスマッチが出たら、サーチ位置を１ずらす。
・各エントリーの２０塩基についてサーチする。
・もしマッチしなかったら、ゲノム配列を反転５′←３′でサーチしなおす。

（ｂ）短いプライマー候補塩基配列と新データベースのエントリーとのホモロジーサーチ



・１個ミスマッチが出たら、サーチ位置を１ずらす。

【図５】


スタート
S21
エントリー評価配列作成・初期化
遺伝的アルゴリズム用２次元配列作成・初期化処理
S22
遺伝的アルゴリズム用２次元配列が各行について
何エントリーと相補するかの計算処理
S23
相補した数の上位から、各行を並べ替える
S24
1．２次元配列の上位１／４（ａ）については、そのまま置く。
2．上位１／４から１／２（ｂ）までについて、上位１／４からランダムで選んだ組み合わせと、別の上位１／４から選んだ組み合わせの各プライマー候補配列のどちらかを乱数で選ぶ。
3．下位１／２の組み合わせ（ｃ）は、完全に乱数で選ぶ。
4．同一のプライマー候補配列を持ったもの、また、同一の組み合わせは除外して、新たに乱数で選ぶ。
S25
数百回以上、繰り返したか？
NO
YES
S26
最適なプライマー候補塩基配列群を上位順に出力
エンド

【図７】

・エントリー評価配列（ａｒｒａｙ）

エントリー＃０　　エントリー＃ｎ

０００００・・・・・・・・・・・・・・・・・・・・・・・０

↑１ビット

・遺伝的アルゴリズム用２次元配列（ａｒｒａｙ）

横列（プライマー候補塩基配列の組み合わせ）

AGCGCT	CGAGCA	AAGTTA	CGCGTT	ATGCTG	AAACCC	TATCTG	CGCCGT
GCTAGC	GGAGGC	TAGCCA	CACAGA	ATGTGG	CAACAA	GAGTAC	CGAACT
ACCTGC	CGAGGA	・	・	・	・	・	・
AACGTC	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・
・	・	・	・	・	・	・	・

縦　行
（100行以上）

【図８】

Primer [6] mer. Total of sequences: 2798

Mean of all sequences: 331.5797

No.	Sequence:	How many matched (Seq. covered):	Mean/Std dev. of loc.:	
0	AAAAAA	2422 (968)	241.6841 /	102.9200
4095	TTTTTT	2186 (836)	239.3683 /	122.7736
819	TATATA	2032 (1077)	266.9060 /	88.3254
3276	ATATAT	1930 (1003)	258.4896 /	92.9654
3072	AAAAAT	1459 (1101)	230.6381 /	105.5594
4030	CTTCTT	1360 (847)	163.0235 /	98.4548
768	AAAATA	1287 (980)	235.7964 /	103.6617
48	AATAAA	1254 (830)	236.5343 /	105.0341
12	ATAAAA	1240 (941)	243.2855 /	102.1883
3	TAAAAA	1225 (934)	237.8808 /	105.4180
192	AAATAA	1220 (884)	237.8787 /	105.0386
4031	TTTCTT	1183 (882)	184.0896 /	104.0615
64	AAAGAA	1167 (885)	219.8955 /	101.7196
4079	TTCTTT	1158 (890)	180.6209 /	104.9007
3055	TTCTTC	1155 (771)	160.4450 /	96.4877
16	AAGAAA	1152 (917)	216.8681 /	103.5133
3840	AAAATT	1092 (869)	217.0421 /	110.0152
4047	TTATTT	1087 (856)	242.9724 /	107.6030
1024	AAAAAG	1079 (850)	223.3411 /	108.7999
4092	ATTTTT	1050 (859)	211.7924 /	116.4093
3264	AAATAT	1040 (861)	238.1173 /	102.2256
2047	TTTTTG	1032 (847)	196.0669 /	113.7874
256	AAAAGA	1012 (817)	228.4733 /	105.6524
4	AGAAAA	999 (827)	224.8398 /	100.7646

【図９】

Primer [6] mer. Total of sequences: 2364

Combination:	How many seq. covered:
TATATA TCCAAA TTCCTT TTTTTT AAGAAG TTTGAA TTATTT TTCAAA	: 2359
ATTTAT CTTCTT GAAGAA TTTTTT TATATA AAAATT TTATTT TTCAAA	: 2358
ATTTAT TCCAAA TTCCTT TTTTTT AAGAAG TTTGAA TTATTT TTCAAA	: 2358
ATTTAT TCCAAA TTCCTT TTTTTT TATATA GAAAGA TTGGAA TTCAAA	: 2358
AATTTT GAAGAA ACAAAG TTTAAT TATATA AAAATT TTAAAT TTCAAA	: 2358
TATATA TCAAAA TTCCTT TTTTTT AAGAAG TTTGAA TTATTT TTCAAA	: 2357
ATTTAT CTTCTT AATTTT TTTTTT AAAATT GAAAGA TTTTCA TTCAAA	: 2357
TATATA TCCAAA AATTTT TTTTTT TGATTT TTTGAA TTATTT TTCAAA	: 2357
AAAATG CTTCTT TTATTT TTTTTT ATTGAA AAAATT CAACAA TTTGAA	: 2357
AAAATG CTTCTT GAAGAA TTTTTT TATATA AAAATT TTATTT TTCAAA	: 2357
AAAATG TCCAAA TTATTT TTTTTT ATTGAA AAAATT CAACAA TTTGAA	: 2357
TATATA TCCAAA TTTATT TTTTTT TTTGGT TTTGAA TCAAAT TCAAAA	: 2357
TTCTTT AAAATT AAATCA TTTTTT TATATA GAAAGA TTTTCA TTCAAA	: 2357
AAAATG TCCAAA CTTTGG TTTTTT TATATA TAATTT TTATTT TTCAAA	: 2357
ATTTAT CTTCTT GAAGAA TTTTTT TATATA ATATTA TTCAAA TTTGAA	: 2357
TTTTTG TGAAGA AAATCA TTTTTT AGAATG AAAATT TTCTTG TTCAAA	: 2357
AAAATG TCCAAA TTCCTT TTCTTT TATATA TTTGAA TTATTT TTCAAA	: 2357
AAAATG TCCAAA TTCAAA TTTTTT TATATA TTTGAA TTATTT TTCTTG	: 2357
AAAATG TCCAAA ATAAAG TTTTTT TATATA TTTGAA TTATTT TTCAAA	: 2357
ATTTAT TGAAGA AATTTT TTTTTT AAAATT GAAAGA TTCAAA TTCTTG	: 2357
ATTTAT CTTCTT GAAGAA TTTTTT TATATA TTTGAA TTATTT TTCAAA	: 2357
AAAATG TCCAAA TTTTTA TTTTTT TATATA TTTGAA TTATTT TTCAAA	: 2357
TTCTTT AAAATT AAATCA TTTTTT AAGAAT TATGCT AATTTT AGAAAG	: 2357
TATATA TCCAAA AATTTT TTTTTT AAAATG TTTGAA TTATTT TTCAAA	: 2357
ATTTAT AAAATT GAAGAA TTTTTT TATATA TTTGAA TTATTT TTCAAA	: 2357
ATTTAT CTTCTT GAAGAA TTTTTT TATATA TTTGAA TTATTT TTCAAA	: 2357

フロントページの続き

(72)発明者 後藤 順
東京都文京区本駒込１－13－４

Ｆターム(参考) 4B024 AA20 BA80 HA11
4B063 QA13 QQ42
5B075 PP22 PQ72 PR04 PR06 QM08
UU19